A genome-wide association study identifies new loci for Factor VII and implicates Factor VII in the etiology of ischemic stroke.

Paul S. de Vries[1,*], Maria Sabater-Lleal[2,*], Jennifer E. Huffman[3,4,5,*], Jonathan Marten[6,*], Ci Song[3,4,7,8,*], Nathan Pankratz[9], Traci Bartz[10], Hugoline G. de Haan[11], Graciela E. Delgado[12], John D. Eicher[3,4], Angel Martinez[13], Cavin K. Ward-Caviness[14], Jennifer Brody[15], Ming-Huei Chen[3,4], Moniek P.M. de Maat[16], Mattias Frånberg[2], Marcus E. Kleber[12,17], Kiang J. Liu[18], Fernando Rivadeneira[19], José Manuel Soria[13], Weihong Tang[20], Geoffrey H. Tofler[21], André G. Uitterlinden[19], Astrid van Hylckama Vlieg[11], Eric Boerwinkle[1,22], Neil M. Davies[23], Oscar H. Franco[24], Barbara McKnight[10], Bruce M. Psaty[25,26], Alex P. Reiner[27,28], INVENT Consortium, MEGASTROKE Consortium, Myriam Fornage[1,29], Anders Hamsten[2], Winfried März[12,30,31], Frits R. Rosendaal[11,32,33], Juan Carlos Souto[34], Abbas Dehghan[35], Andrew D. Johnson[3,4,‡], Alanna C. Morrison[1,‡], Christopher J. O'Donnell[3,36,‡], and Nicholas L. Smith[26,28,37,‡]

[*]Contributed equally to this manuscript as junior authors
[‡]Contributed equally to this manuscript as senior authors

[1]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA
[2]Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
[3]Population Sciences Branch, National Heart, Lung, and Blood Institute, Framingham, MA, USA
[4]The Framingham Heart Study, Framingham, MA, USA
[5]Center for Population Genomics, VA Boston Healthcare System, Jamaica Plain, MA, USA
[6]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland
[7]Department of Medical Sciences, Uppsala University, Uppsala, Sweden
[8]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
[9]Department of Laboratory Medicine and Pathology, University of Minnesota School of Medicine, Minneapolis, MN, USA
[10]Department of Biostatistics, University of Washington, Seattle, WA, USA
[11]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[12]Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[13]Unitat de Genómica de Malalties Complexes, Institut d'Investigació Biomedica Sant Pau (IIB-Sant Pau), Barcelona, Spain
[14]National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, Chapel Hill, NC, USA
[15]Department of Medicine, University of Washington, Seattle, WA, USA
[16]Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands
[17]Institute of Nutrition, Friedrich-Schiller-University Jena, Mannheim, Germany
[18]Department of Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA
[19]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands
[20]Division of Epidemiology & Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA
[21]Royal North Shore Hospital, University of Sydney, Sydney, Australia
[22]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

[23]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
[24]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
[25]Department of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA
[26]Kaiser Permanente Washington Research Institute, Kaiser Permanente Washington, Seattle, WA, USA
[27]Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[28]Department of Epidemiology, University of Washington, Seattle, WA, USA
[29]Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA
[30]Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
[31]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University Graz, Graz, Austria
[32]Einthoven Laboratory of Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands
[33]Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden, the Netherlands
[34]Unitat d'Hemostasia i Trombosi, Hospital de la Sant Creu i Sant Pau, Barcelona, Spain
[35]Department of Epidemiology and Biostatistics, Imperial College London, London, UK
[36]Cardiology Section, VA Boston Healthcare System, West Roxbury, MA, USA
[37]Seattle Epidemiologic Research and Information Center, Department of Veteran Affairs, Office of Research and Development, Seattle, WA, USA

Correspondence

Nicholas L. Smith, PhD

Department of Epidemiology, University of Washington

1730 Minor Ave, Suite 1360, Seattle, WA, USA, 98101

Phone: +1 206 221 7775

Email: nlsmith@u.washington.edu

Abstract word count: 248

Main text word count: 3,944 including acknowledgements, funding, and disclosures

Number of Tables: 2

Number of Figures: 4

Number of Supplementary Tables: 4

Number of Supplementary Figures: 11

Abstract ***(n=248 words, limit=250)***

Factor VII (FVII) is an important component of the coagulation cascade. Few genetic loci regulating FVII activity and/or levels have been discovered to date.

We conducted a meta-analysis of nine genome-wide association studies of plasma FVII levels (seven FVII activity and two FVII antigen) among 27,495 participants of European and African ancestry. Each study performed ancestry-specific association analyses. Inverse variance weighted meta-analysis was performed within each ancestry group and then combined for a trans-ancestry meta-analysis. Our primary analysis included the seven studies that measured FVII activity, and a secondary analysis included all nine studies. We provided functional genomic validation for newly identified significant loci by silencing candidate genes in a human liver cell line (HuH7) using siRNA and then measuring *F7* mRNA and FVII protein expression. Lastly, we used meta-analysis results to perform Mendelian randomization analysis to estimate the causal effect of FVII activity on coronary artery disease, ischemic stroke, and venous thromboembolism.

We identified two novel (*REEP3* and *JAZF1-AS1*) and six known loci associated with FVII activity, explaining 19.0% of the variance. Adding FVII antigen data to the meta-analysis did not result in the discovery of further loci. Silencing *REEP3* in HuH7 cells upregulated FVII, while silencing *JAZF1* downregulated FVII. Mendelian randomization analyses suggest that FVII activity has a positive causal effect on the risk of ischemic stroke.

Variants at *REEP3* and *JAZF1* contribute to FVII activity by regulating *F7* expression levels. FVII activity appears to contribute to the etiology of ischemic stroke in the general population.

Introduction

As the initiator of the extrinsic coagulation pathway, coagulation factor VII (FVII) plays a central role in fibrin formation. FVII and tissue factor activate factor X, which then converts prothrombin to thrombin, and in turn converts fibrinogen into fibrin. Plasma levels of FVII are associated with several clinical outcomes. For example, FVII deficiency is a rare bleeding disorder associated with hemorrhagic complications,[1] while elevated levels of FVII have been associated with arterial thrombosis and venous thromboembolism (VTE).[2-5]

FVII activity and levels have a substantial heritable component, with estimates of the heritability of FVII activity ranging from 0.40 to 0.52.[6,7] The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium previously conducted genome-wide association studies (GWASs) with data on over 2 million common single nucleotide polymorphisms (SNPs) in European-ancestry participants, identifying four new candidate genes for FVII in addition to the protein-coding locus, *F7*.[8,9] The lead variants at these known loci explain 7.7% of the variance in FVII, implying that further heritability remains to be uncovered.[8]

To discover additional genetic variants associated with FVII, we performed an expanded GWAS with data on over 10 million common and low-frequency SNPs and insertion-deletions in 27,495 participants across nine studies, including 3,420 African American participants. Gene silencing in a human liver cell line was used to validate the genomic function of significantly associated loci. Lastly, we performed Mendelian randomization analyses to estimate the causal effects of FVII activity on atherosclerotic and thrombotic diseases by leveraging our GWAS results in Mendelian randomization analyses.

Methods

Study design and Participating Cohorts

This study was organized within the CHARGE Consortium Hemostasis Working Group.[9] Nine studies were included: the Atherosclerosis Risk in Communities (ARIC) study,[10] the Cardiovascular Health Study (CHS),[11] the Coronary Artery Risk Development In young Adults (CARDIA) study,[12] the Genetic Analysis for Idiopathic Thrombophilia 2 (GAIT2) study, the Framingham Heart Study

(FHS),[13] the LUdwigshafen RIsk and Cardiovascular Health (LURIC) study,[14] the Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA) study, the Precocious Coronary Artery DISease (PROCARDIS) study,[15] and the Rotterdam Study (RS).[16] Descriptions and ancestry composition of participating cohorts are found in the Supplementary Material (**Supplementary Methods** and **Supplementary Table 1**). Seven studies (ARIC, CHS, CARDIA, GAIT2, LURIC, MEGA, RS) including 23,434 participants measured FVII activity (% or IU/ml*100) and two studies (FHS, PROCARDIS) including 4,061 participants measured FVII antigen (% or IU/ml*100).

Genotyping and imputation

All participating cohorts performed genome-wide genotyping using commercial platforms available from Illumina or Affymetrix. Each study performed standard pre-imputation quality control checks and imputed autosomal and X-chromosome variants from the 1000 Genomes Project (1000G) Phase I version 3 reference panel using available imputation methods.[17-20] Genotyping, pre-imputation quality control, and imputation procedures are described in **Supplementary Table 2**.

Cohort-specific association analyses

Natural-log-transformed FVII was analyzed within each cohort. Participants with values 3 standard deviations above or below the population mean were removed prior to cohort-level analysis and any individuals on anticoagulant therapy were also excluded. Ancestry-stratified, study-specific regression analyses using an additive genetic model were performed between genome-wide 1000G imputed variant dosages and phenotype levels, adjusted for age, sex, ancestry-informative principal components, and study-specific variables, such as center. Analyses of the X-chromosome were stratified by sex, with variants in males coded as 0/2. The covariates used in each of the studies are shown in **Supplementary Table 2**. Quality control assessment of ancestry-specific results files from each study was conducted prior to meta-analysis using the EasyQC software package.[21] Quality control procedures are further described in the **Supplementary Methods**.

Trans-ancestry meta-analysis

The discovery trans-ancestry meta-analysis was conducted in two steps. First, METAL was used to perform ancestry-specific inverse-variance weighted meta-analysis.[22] We then used the same method to meta-analyze the ancestry-specific results. As suggested by Huang et al.[23], we adopted a genome-wide significance threshold of *P*-value < 2.5×10^{-8}. Compared with the traditional genome-wide significance threshold of 5×10^{-8}, this stricter threshold additionally corrects for the low-frequency variants that were not included in the initial generation of GWASs.[24] Finally, a locus was defined as +/- 1Mb from the variant with the lowest *P*-value.

In order to reduce heterogeneity the primary trans-ancestry meta-analysis included the seven studies that measured FVII activity, and not the two studies (FHS and PROCARDIS) that measured FVII antigen. In a secondary meta-analysis we added results from the two studies that measured FVII antigen.

Post-discovery analyses

Newly identified loci were validated and characterized by using siRNA to silence candidate genes in human liver HuH7 cells, and measuring F7 mRNA levels and release of FVII protein levels. These functional validation steps are described in detail in the **Supplementary Methods**.

To identify additional independent signals at the associated loci, an approximate method implemented in GCTA was used for conditional and joint analysis using meta-analysis summary statistics from the trans-ancestry meta-analysis of FVII activity.[25,26] Further details on the conditional analysis are provided in the **Supplementary Methods**.

Mendelian randomization analyses were used to investigate the causal effect of FVII activity on coronary artery disease (CAD), ischemic stroke (IS), and VTE. We used two-sample methods that rely on summary statistics (beta coefficients with standard errors from GWASs).[27] We obtained summary statistics for CAD from the CARDIoGRAMplusC4D consortium (http://www.cardiogramplusc4d.org/data-downloads/),[28] summary statistics for IS from the MEGASTROKE consortium,[29] and summary statistics for VTE from the INVENT consortium.[30] The methods used to perform Mendelian randomization can be found in the **Supplementary Methods**. In

brief, we used four techniques to obtain causal effect estimates based on the lead variants at the genome-wide significant loci: 1) inverse-variance weighted meta-analysis (primary analysis), 2) Egger regression,[31] and 3) weighted median estimator,[32] 4) restriction of the analysis to the lead variant at the *F7* locus. Given that the lead variant at the *F7* locus is located in the gene that encodes the FVII protein, it may be less likely to influence clinical outcomes through pathways that do not involve FVII.

Results

Baseline characteristics

In total, 20,014 European-ancestry and 3,420 African-ancestry subjects from seven studies were included in the meta-analysis of FVII activity and an additional 4,061 European-ancestry subjects were included in the combined meta-analysis of FVII activity and antigen. Baseline characteristics are shown in **Supplementary Table 1**. The mean age across the studies was 57.2 years, and 52.2% of the participants were women.

Trans-ancestry meta-analysis

After quality control, 10,044,948 variants across the autosomal and X chromosomes were examined in the trans-ancestry meta-analysis of FVII activity. Of these variants 9,316,598 were SNPs and 728,350 were insertions-deletions. The genomic inflation factors that were used to apply genomic control correction to each of the included studies were all < 1.05 and are shown in **Supplementary Table 2**. A QQ plot and Manhattan plot are shown in **Supplementary Figures 1 and 2**, respectively.

Genome-wide significant results are presented in **Table 1**. Briefly, 1,637 variants located in eight loci exceeded the genome-wide significance level of *P*-value < 2.5×10^{-8}. Among the associated regions, loci containing *F7, PROC, GCKR, MS4A6A, ADH4,* and *TSKU* represented replications of previously described loci (**Supplementary Figures 3-8**),[8] whereas two loci were novel: *REEP3* and *JAZF1-AS1*. The most significant variant at the *REEP3* locus was an intronic variant, rs10761784 (Beta = 0.013; *P*-value = 6.7×10^{-10}) in *REEP3* (**Figure 1**). At the second novel locus the lead variant, rs498475, was located within the noncoding RNA *JAZF1-AS1* (Beta = 0.012; *P*-value = 1.5×10^{-8};

Figure 2). Lead variants at PROCR and GCKR were identical to previously reported lead variants, whereas the lead variants at the remaining known loci differed from previously reported lead variants (**Supplementary Table 3**).

No additional genome-wide significant loci emerged when adding data from two additional studies in the combined analysis of FVII activity and antigen, but variants at the *TSKU* and *JAZF1-AS1* loci were no longer genome-wide significant (**Supplementary Table 4**). A QQ plot and Manhattan plot for the combined analysis of FVII activity and antigen are shown in **Supplementary Figures 9 and 10** respectively. The lead variants at TSKU and JAZF1-AS1 had opposing effect directions on FVII activity and antigen, but lead variants at the remaining six loci had relatively similar effects on FVII activity and antigen (**Supplementary Figure 11**). The variance in FVII activity explained by the lead variants at the eight significant loci was 17.6%. The variance explained by each of the lead variants individually is shown in **Table 1**.

Conditional analysis

Conditional analysis identified four independent signals at the *F7* locus as well as two independent signals at the *PROCR* locus. The conditional analysis of the trans-ancestry meta-analysis of FVII activity is shown in **Table 2**. Among the independently associated variants at the *F7* locus was a low-frequency variant (minor allele frequency = 0.02) with the second largest effect size discovered by GWASs thus far (joint beta = 0.08; joint *P*-value = 8.7×10^{-20}). By considering these independent signals, the variance in FVII activity explained by the *F7* locus increased from 13.9% to 15.2%, while the variance explained by the *PROCR* locus increased from 1.6 to 1.8%. The total variance explained therefore increased from 17.6% to 19.0%.

Functional validation of novel loci

The s47939 and s37271 silencing siRNAs both suppressed expression of *REEP3* mRNA by 88% compared with the scramble siRNA (negative control). Experiments were repeated three times with consistent results, showing that silencing of *REEP3* resulted in upregulation of *F7* mRNA (*P*-value =

0.0001 for s47939; *P*-value > 0.05 for s37271; **Figure 3**) and a corresponding increase in FVII protein levels (*P*-value = 9.1×10^{-5} for s47939; *P*-value = 0.0003 for s37271; **Figure 3**).

At the *JAZF1-AS1* locus we targeted the *JAZF1* gene for silencing rather than the antisense noncoding RNA in which the lead variant was located. The s225897 silencer reduced expression of *JAZF1* mRNA by 68%, whereas the s48121 silencer reduced *JAZF1* mRNA expression by 75%. As shown in **Figure 3**, silencing of *JAZF1* resulted in downregulation *F7* mRNA (*P*-value = 0.02 for s225897; *P*-value < 2×10^{-6} for s48121) and a corresponding decrease in FVII protein expression: silencing experiments showed no effect on FVII protein in the media of cells silenced with s225897, but a significant decrease upon silencing with s48121 (*P*-value = 1.1×10^{-6}).

Mendelian randomization

Figure 4 contains forest plots showing causal effect estimates of FVII activity on A) CAD, B) IS, and C) VTE. Causal effect estimates are given as odds ratios (ORs) per 1 unit increase in natural-log-transformed FVII activity (% or IU/ml*100). Effect estimates were obtained from single variants associated with FVII activity and meta-analyzed to produce combined causal effect estimates. Heterogeneity was detected among the causal effect estimates ($P_{heterogeneity}<0.05$), and the variant at the *PROCR* locus, rs867186, was removed from all analyses due to outlying causal effect estimates for CAD, IS, and VTE. No further heterogeneity was detected after excluding this variant ($P_{heterogeneity}>0.05$).

A significant causal effect of FVII activity on IS was detected (OR_{IVW} = 1.37; 95% confidence interval (CI95) = 1.14-1.65). Given that the SD of natural-log-transformed FVII activity ranged from 0.18 to 0.26 across our studies, the causal effect estimate corresponds to an approximate OR of 1.06 to 1.09 per SD change in natural-log-transformed FVII activity. Results were consistent across sensitivity analyses, including the use of Egger regression, the weighted median estimator, and restriction of the analysis to the rs561241 variant at the *F7* locus, indicating that pleiotropy is unlikely to have biased the causal estimate. Causal effect estimates for CAD (OR_{IVW} = 1.14; CI95 = 0.97-1.34) and VTE (OR_{IVW} = 1.22; CI95 = 0.80-1.85) were more modest and failed to reach statistical

significance. Nevertheless, the magnitude of these effect estimates were consistent across sensitivity analyses, including when the rs561241 variant at the *F7* locus was examined in isolation (OR_{CAD} = 1.14; CI95 = 0.96-1.36; OR_{VTE} = 1.31; CI95 = 0.84-2.07).

Discussion

In this GWAS of circulating FVII levels, we identified the six previously known FVII loci as well as two new loci: *REEP3* and *JAZF1-AS1*. In total, the eight loci associated with FVII activity explained 19.0% of the variance. For each new discovery, we showed functional impact *in vitro* of candidate genes on *F7* mRNA and FVI protein expression: *REEP3* gene silencing in liver cells increased *F7* mRNA and FVII protein expression, whereas *JAZF1* gene silencing decreased *F7* mRNA and FVII protein expression.

REEP3 encodes Receptor Accessory Protein 3. Although this protein has not been widely studied, there is evidence that an absence of this protein leads to defects in mitosis and a proliferation of intranuclear membranes derived from the nuclear envelope.[33] The *REEP* gene family may also be involved in shaping the membrane of the endoplasmic reticulum and the trafficking of G-protein coupled receptors.[34] Given that FVII is processed in the endoplasmic reticulum, this may explain the association with FVII levels.[35] The locus containing *REEP3* has been previously associated to several other coagulation phenotypes, namely circulating fibrinogen levels,[36,37] mean platelet volume,[38,39] and platelet aggregation,[40] as well as to liver enzyme concentrations.[41,42] For many of these phenotypes, the gene that was reported is not *REEP3* but *JMJD1C*, with missense variants localized in the *JMJD1C* being associated with mean platelet volume.[38] Functional studies in zebrafish indicate that *JMJD1C* has a major role in hematopoiesis.[43] Although we did not examine the consequences of *JMJD1C* silencing on *F7* expression and FVII release, our experiments implicate *REEP3* as a causal gene for FVII. These results are consistent with tissue-specific pleiotropic effects at this locus, with *JMJD1C* being involved in hematopoiesis and *REEP3* being of functional relevance in the liver, although further research is needed to confirm this hypothesis.

JAZF1-AS1 is a non-coding RNA that may regulate the adjacent *JAZF1* gene, which encodes a transcriptional repressor. Variants at the *JAZF1-AS1* locus were associated with FVII activity, but

their effect on FVII was attenuated when we included studies that measured FVII antigen. A possible explanation is that variants at the *JAZF1-AS1* loci affect FVII activity independent of circulating FVII protein levels. However, silencing of *JAZF1* in liver cells resulted in lower *F7* mRNA and FVII protein expression, suggesting that the mechanism underlying the genetic association is likely to involve FVII levels.

Apart from *REEP3* and *JAZF1-AS1* we identified six known loci: *F7*, *PROCR*, *GCKR*, *MS4A6A*, *ADH4*, and *TSKU*. The results of this study may aid in the identification of causal variants at these loci. For example, lead variants in *PROCR* and *GCKR* were both missense variants leading to amino acid substitutions (Ser219Gly in *PROCR* and Pro446Leu in *GCKR*). These variants were also the lead variants in their respective loci in the previous GWAS of FVII,[8] and have been associated with other hemostatic phenotypes.[44,45] In contrast, the lead variants that we identified at the *F7*, *MS4A6A*, *ADH4*, and *TSKU* loci differ from those published in the previous GWAS and may be in higher linkage disequilibrium with the true causal variant.[8]

Using the genetic association results generated in this study, we performed Mendelian randomization analyses to estimate the causal effect of FVII activity on CAD, IS, and VTE. These analyses suggest that variation in FVII activity in the general population influences the risk of IS. These results warrant further etiological research on the role of FVII in IS, as well as translational research on potential clinical applications involving FVII. Potential clinical applications that should be investigated include the reduction of FVII activity through lifestyle or pharmaceutical interventions for the prevention of IS, as well as the restriction of off-label use of recombinant FVII given the increased risk of IS that this may result in.[46-48]

Our results do not exclude the possibility of additional causal effects of FVII activity on CAD and VTE. In fact, when using the rs561241 variant at the *F7* locus in isolation as an instrumental variable, the estimate of the causal effect on VTE was equal to the estimate of the effect on IS. Lower statistical power for the Mendelian randomization analysis of VTE may explain the lack of a statistically significant causal effect of FVII activity on this outcome: the GWAS from which we obtained the effect of the variants on IS was comprised of 60,341 cases and 454,450 controls,[29] while

the GWAS on VTE consisted of 7,507 cases and 52,632 controls.[30] Our results are thus consistent with causal effects of FVII activity on CAD and VTE, albeit more modest effects than on IS. Further research using larger sample sizes will be necessary to detect or rule out these effects.

Our GWAS included 27,495 participants, providing a 78% increased sample size when compared with the largest previous GWAS of FVII levels.[8] Other strengths include the functional validation of newly identified loci by silencing candidate genes in human liver cell lines, as well as the use of Mendelian randomization in leveraging our newly generated data for insights into disease etiology. However, these approaches also have limitations. In the gene silencing experiments, we silenced a single gene at each locus. Because an effect of FVII levels was observed in both cases, we did not pursue further experiments involving other genes at these loci. As such we cannot exclude the possibility that other genes at these loci also influence FVII. In turn, the validity of Mendelian randomization can be threatened by the presence of pleiotropic effects among FVII variants. In order to minimize the impact of pleiotropy on our results, we excluded variants with heterogeneous effects on disease outcomes from the analyses. This lead to the removal of the variant at the *PROCR* locus from all analyses. Furthermore, we performed three alternative Mendelian randomization approaches, Egger regression, weighted median estimation, and restricting the analysis to the rs561241 variant at the *F7* locus, that are, to some degree, robust to pleiotropy.[31,32] The estimate of the causal effect of FVII activity on IS was consistent across these sensitivity analyses, as were the estimates of the causal effects of FVII activity on CAD and VTE.

To conclude, this study identifies two novel loci associated with FVII activity and functional studies suggest that *REEP3* and *JAZF1* are the causal genes within these loci. Mendelian randomization analyses indicate that FVII activity is causally involved in the development of IS and possibly CAD and VTE, with high FVII activity being associated with an increased risk of these clinical outcomes.

Acknowledgements

We thank the staff and participants of each of the participating studies. We are grateful to Alexandra Bäcklund, Fariba Foroogh and Angela Silveira for helpful advice in the cell culture and mesoscale experiments.

Funding

This work was supported by American Heart Association Grant #17POST33350042 (Paul S. de Vries) and National Heart, Lung, and Blood Institute (NHLBI) grant R01HL134894 (Nicholas L. Smith). Maria Sabater-Lleal is recipient of the EHA-ISTH fellowship and is partially supported by the Swedish Heart-Lung Foundation (20160290). Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756.

The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by the NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

Cardiovascular Health Study (CHS) research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL085251, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and R01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in

part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging NIA) and an intra-agency agreement between NIA and NHLBI (AG0005). Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed and approved by CARDIA for scientific content.

Framingham Heart Study (FHS) was partially supported by the National Heart, Lung, and Blood Institute's (NHLBI's) Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. JEH, CS, MHC, ADJ and CJO were supported by National Heart, Lung and Blood Intramural Research Funds. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U. S. Department of Health and Human Services.

Genetic Analysis for Idiopathic Thrombophilia 2 (GAIT2) project was supported partially by grants PI-11/0184, PI-14/0582 and Red Investigación Cardiovascular RD12/0042/0032 from the Instituto

Carlos III (Fondo de Investigación Sanitaria–FIS), and 2014SGR-402 Consolidated Research Group of the Generalitat de Catalunya.

The LUdwigshafen RIsk and Cardiovascular Health (LURIC) study was supported by the 7th Framework Program RiskyCAD (grant agreement number 305739) of the European Union. The work of W.M. and M.E.K. is supported as part of the Competence Cluster of Nutrition and Cardiovascular Health (nutriCARD) which is funded by the German Federal Ministry of Education and Research.

The Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA) study was supported by Netherlands Heart Foundation (NHS 98.113), the Dutch Cancer Foundation (RUL 99/1992), the Netherlands Organisation for Scientific Research (912–03–033| 2003), and partially by the GenMed LABEX (ANR-10-LABX-0013). We would like to thank all colleagues from the French Centre National de Génotypage for the genotyping contribution.

The Precocious Coronary Artery DISease (PROCARDIS) study project was funded with 10 million Euros through the 6th Framework Program of the European Union (LSH-2005-2.1.1-1). It started in Apr 2007 and funding lasted until Sept 2011. After the funding ended, PROCARDIS is still actively doing research in the field of coronary artery disease genetics.

The generation and management of GWAS genotype data for the Rotterdam Study (RS) is supported by the Netherlands Organisation of Scientific Research NWO Investments (no. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project no. 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

Disclosures

Winfried März reports grants and personal fees from AMGEN, BASF, Sanofi, Siemens Diagnostics, Aegerion Pharmaceuticals, Astrazeneca, Danone Research, Numares, Pfizer, Hoffmann LaRoche: personal fees from MSD, Alexion; grants from Abbott Diagnostics, all outside the submitted work. Winfried März is employed by Synlab Holding Deutschland GmbH.

References

1. Lapecorella M, Mariani G, International Registry on Congenital Factor VIID. Factor VII deficiency: defining the clinical picture and optimizing therapeutic options. *Haemophilia*. 2008;14(6):1170-1175.
2. Zakai NA, Lange L, Longstreth WT, Jr., et al. Association of coagulation-related and inflammation-related genes and factor VIIc levels with stroke: the Cardiovascular Health Study. *J Thromb Haemost*. 2011;9(2):267-274.
3. Smith A, Patterson C, Yarnell J, Rumley A, Ben-Shlomo Y, Lowe G. Which hemostatic markers add to the predictive value of conventional risk factors for coronary heart disease and ischemic stroke? The Caerphilly Study. *Circulation*. 2005;112(20):3080-3087.
4. Tsai AW, Cushman M, Rosamond WD, et al. Coagulation factors, inflammation markers, and venous thromboembolism: the longitudinal investigation of thromboembolism etiology (LITE). *Am J Med*. 2002;113(8):636-642.
5. Folsom AR. Hemostatic risk factors for atherothrombotic disease: an epidemiologic view. *Thromb Haemost*. 2001;86(1):366-373.
6. Freeman MS, Mansfield MW, Barrett JH, Grant PJ. Genetic contribution to circulating levels of hemostatic factors in healthy families with effects of known genetic polymorphisms on heritability. *Arterioscler Thromb Vasc Biol*. 2002;22(3):506-510.
7. Souto JC, Almasy L, Borrell M, et al. Genetic determinants of hemostasis phenotypes in Spanish families. *Circulation*. 2000;101(13):1546-1551.
8. Smith NL, Chen MH, Dehghan A, et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation*. 2010;121(12):1382-1392.
9. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2(1):73-80.
10. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702.
11. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1(3):263-276.
12. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41(11):1105-1116.
13. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med*. 1975;4(4):518-525.
14. Winkelmann BR, Marz W, Boehm BO, et al. Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics*. 2001;2(1 Suppl 1):S1-73.
15. Barlera S, Chiodini BD, Franzosi MG, Tognoni G. [PROCARDIS: A current approach to the study of the genetics of myocardial infarct]. *Ital Heart J Suppl*. 2001;2(9):997-1004.
16. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*. 2015;30(8):661-708.
17. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
18. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.

19. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816-834.
20. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.
21. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-1212.
22. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.
23. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur J Hum Genet*. 2012;20(7):801-805.
24. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet*. 2016;24(8):1202-1205.
25. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369-375, S361-363.
26. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82.
27. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665.
28. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130.
29. Malik R, Chauhan G, Traylor M, et al. Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Submitted*. 2017.
30. Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet*. 2015;96(4):532-542.
31. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525.
32. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304-314.
33. Schlaitz AL, Thompson J, Wong CC, Yates JR, 3rd, Heald R. REEP3/4 ensure endoplasmic reticulum clearance from metaphase chromatin and proper nuclear envelope architecture. *Dev Cell*. 2013;26(3):315-323.
34. Bjork S, Hurt CM, Ho VK, Angelotti T. REEPs are membrane shaping adapter proteins that modulate specific g protein-coupled receptor trafficking by affecting ER cargo capacity. *PLoS One*. 2013;8(10):e76366.
35. Bolt G, Kristensen C, Steenstrup TD. More than one intracellular processing bottleneck delays the secretion of coagulation factor VII. *Thromb Haemost*. 2008;100(2):204-210.
36. de Vries PS, Chasman DI, Sabater-Lleal M, et al. A meta-analysis of 120 246 individuals identifies 18 new loci for fibrinogen concentration. *Hum Mol Genet*. 2016;25(2):358-370.
37. Sabater-Lleal M, Huang J, Chasman D, et al. Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation*. 2013;128(12):1310-1324.
38. Eicher JD, Chami N, Kacprowski T, et al. Platelet-Related Variants Identified by Exomechip Meta-analysis in 157,293 Individuals. *Am J Hum Genet*. 2016;99(1):40-55.
39. Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet*. 2009;41(11):1182-1190.
40. Johnson AD, Yanek LR, Chen MH, et al. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. *Nat Genet*. 2010;42(7):608-613.
41. Chambers JC, Zhang W, Sehmi J, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet*. 2011;43(11):1131-1138.

42. D'Cruz OJ, Haas GG, Jr., de La Rocha R, Lambert H. Occurrence of serum antisperm antibodies in patients with cystic fibrosis. *Fertil Steril*. 1991;56(3):519-527.
43. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*. 2011;480(7376):201-208.
44. Sennblad B, Basu S, Mazur J, et al. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet*. 2017;26(3):637-649.
45. Tang W, Basu S, Kong X, et al. Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood*. 2010;116(23):5032-5036.
46. Logan AC, Yank V, Stafford RS. Off-label use of recombinant factor VIIa in U.S. hospitals: analysis of hospital records. *Ann Intern Med*. 2011;154(8):516-522.
47. Yank V, Tuohy CV, Logan AC, et al. Systematic review: benefits and harms of in-hospital use of recombinant factor VIIa for off-label indications. *Ann Intern Med*. 2011;154(8):529-540.
48. Levi M, Levy JH, Andersen HF, Truloff D. Safety of recombinant activated factor VII in randomized clinical trials. *N Engl J Med*. 2010;363(19):1791-1800.

Tables

Table 1: Lead variants at additional loci associated with FVII activity when excluding studies that measured FVII antigen from the trans-ancestry meta-analysis.

Variant rsID	**Chr:Pos**	**Alleles**	**Freq**	**β**	**Standard Error**	***P*-value**	**Variance Explained**	**Closest Gene**	**Annotation**	**Status**
rs569557	13:113769917	G/A	0.89	0.157	0.003	6.4×10^{-600}	13.9%	*F7*	Intronic	Known
rs867186	20:33764554	G/A	0.10	0.057	0.003	3.3×10^{-64}	1.6%	*PROCR*	Missense	Known
rs1260326	2:27730940	T/C	0.39	0.024	0.002	2.3×10^{-30}	0.7%	*GCKR*	Missense	Known
rs7935829	11:59942815	G/A	0.39	0.018	0.002	6.3×10^{-18}	0.4%	*MS4A6A*	Intronic	Known
rs6532796	4:100042242	G/A	0.70	0.016	0.002	2.6×10^{-13}	0.3%	*ADH4*	Downstream	Known
rs1149616	11:76498369	T/C	0.17	0.017	0.003	1.7×10^{-10}	0.2%	*TSKU*	Intronic	Known
rs10761784	10:65308750	A/T	0.53	0.013	0.002	6.7×10^{-10}	0.2%	*REEP3*	Intronic	Novel
rs498475	7:28256240	G/A	0.42	0.012	0.002	1.5×10^{-8}	0.2%	*JAZF1-AS1*	ncRNA	Novel

The Chr:Pos column shows the chromosome and position (Build 37). The Alleles column shows the FVII-increasing allele / FVII-decreasing allele. The Freq column shows the frequency of the FVII-increasing allele. The variance explained shown in this table was calculated using the sample size weighted mean variance of log-transformed FVII and the betas and frequencies from the trans-ancestry meta-analysis summary statistics.

Table 2: Conditional analysis results for FVII activity using the trans-ancestry meta-analysis results.

rsID	Chr:Pos	Alleles	Freq	β	*P*-value	Joint β	Joint *P*-value	Variance Explained
F7								
rs117989138	13:113697671	A/G	0.02	0.086	3.6×10^{-22}	0.081	8.7×10^{-20}	0.6%
rs36086577	13:113728498	C/A	0.87	0.035	2.2×10^{-19}	0.031	7.5×10^{-15}	0.6%
rs71446935	13:113734376	A/G	0.31	0.035	5.5×10^{-38}	0.032	5.1×10^{-31}	1.2%
rs1046205	13:113752057	A/T	0.79	0.121	3.9×10^{-573}	0.121	$<1.0\times10^{-320}$	12.1%
PROCR								
rs6119569	20:33672371	G/A	0.78	0.022	8.8×10^{-17}	0.019	3.9×10^{-13}	0.3%
rs867186	20:33764554	G/A	0.10	0.057	3.3×10^{-64}	0.055	8.1×10^{-59}	1.4%

The Chr:Pos column shows the chromosome and position (Build 37). The Alleles column shows the FVII-increasing allele. The Freq column shows the frequency of the FVII-increasing allele. The variance explained shown in this table was calculated using the sample size weighted mean variance of log-transformed FVII and the betas and frequencies from the trans-ancestry meta-analysis summary statistics. The β and *P*-value columns are based on the association of each variant in isolation, while the Joint β and Joint *P*-value columns are based on the association of each variant test conditioned on the other variants. The Variance Explained column is based on the joint analysis.

Figures

Figure 1: Regional FVII association plot of the *REEP3* locus in the trans-ancestry meta-analysis of FVII activity.

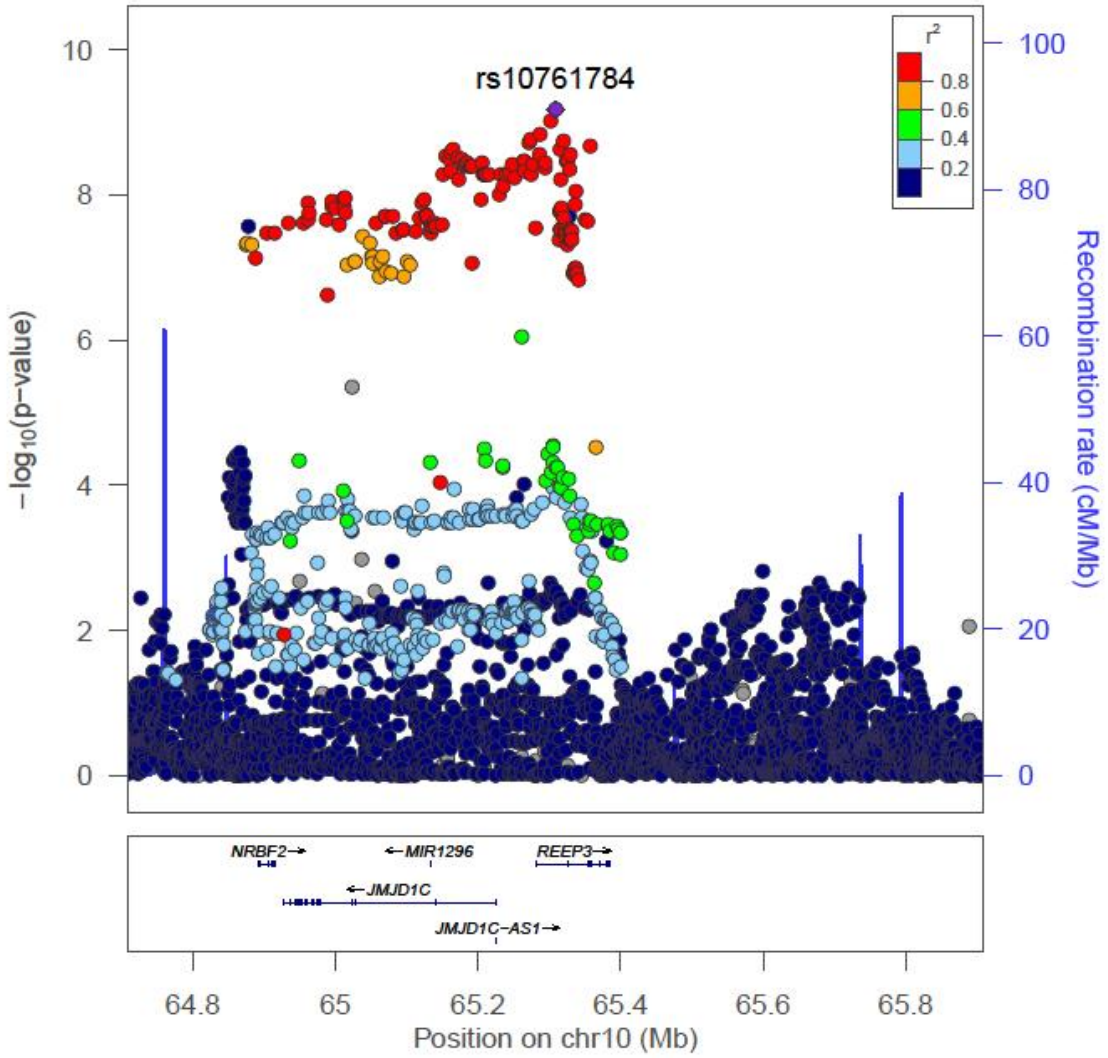


Figure 2: Regional association plot for the *JAZF1* locus in the trans-ancestry meta-analysis of FVII activity.

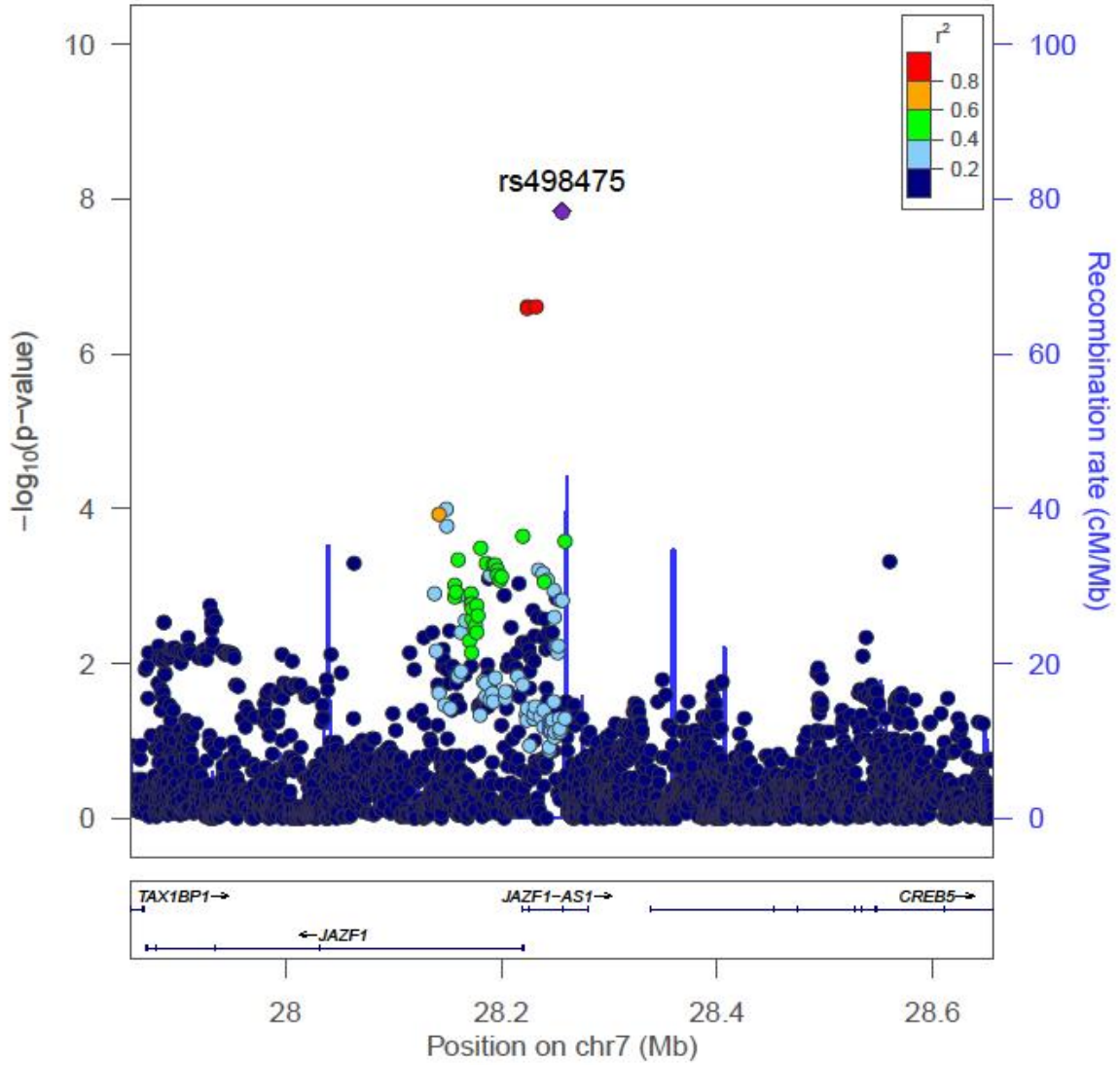


Figure 3: a) *F7* RNA expression after silencing *REEP3*, b) FVII protein levels in cell media after silencing *REEP3*, c) *F7* RNA expression after silencing *JAZF1*, d) FVII protein levels in cell media after silencing *JAZF1*.

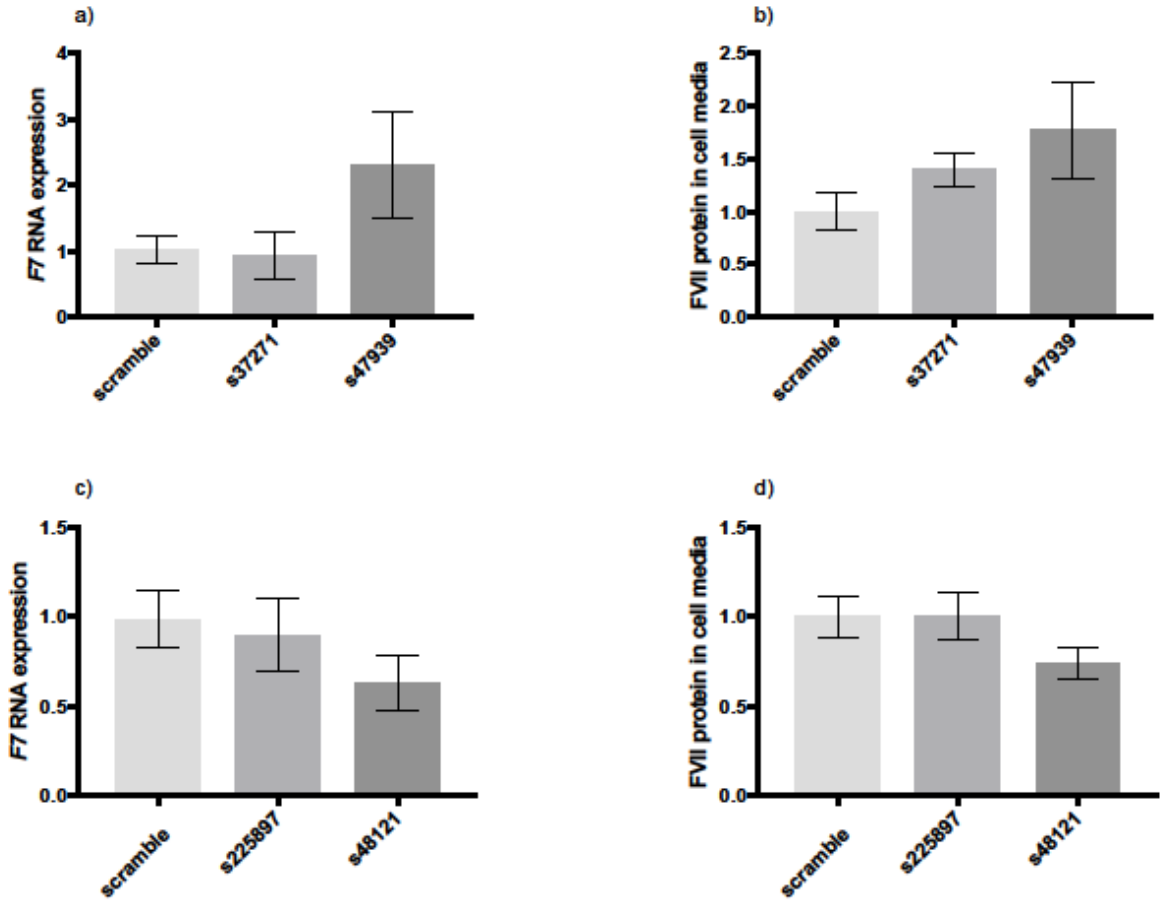
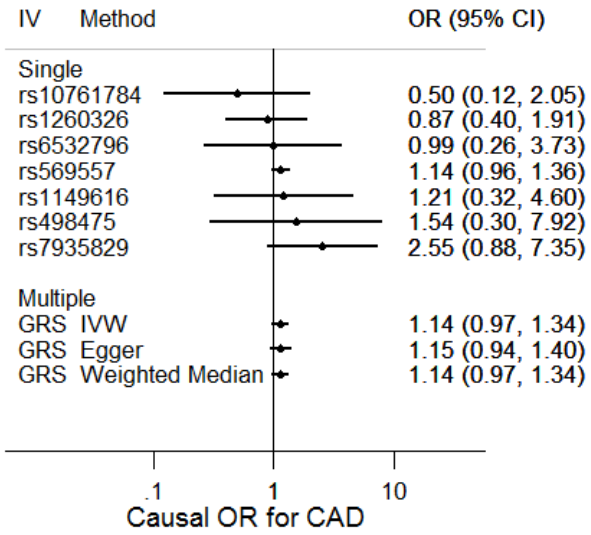
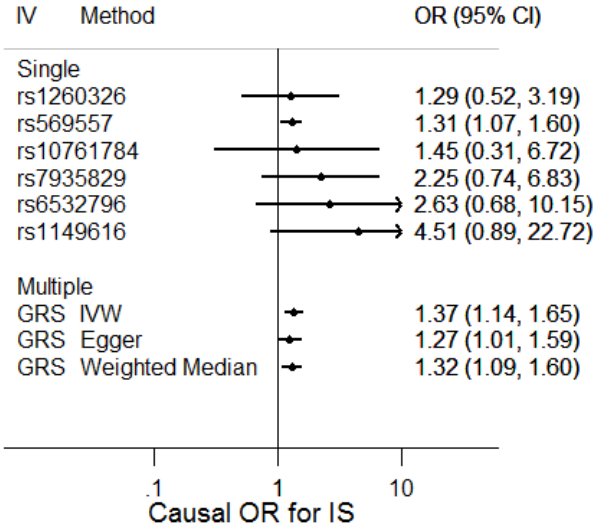
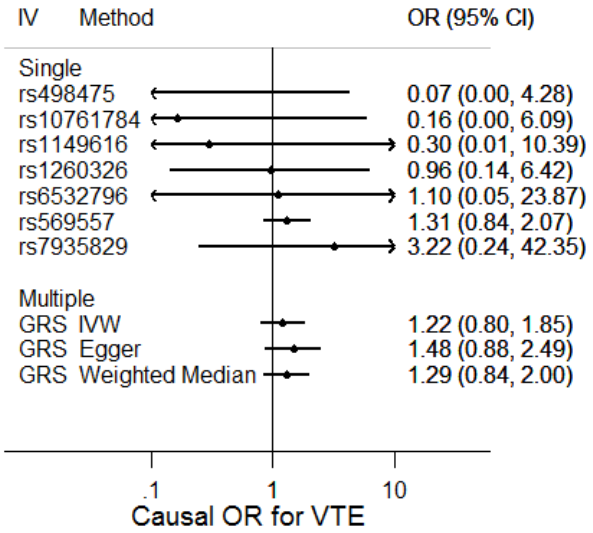


Figure 4: Causal effect estimates of FVII activity on coronary artery disease (CAD), ischemic stroke (IS), and venous thromboembolism (VTE) using Mendelian randomization.


IV Method
OR (95% CI)
Single
rs10761784
0.50 (0.12, 2.05)
rs1260326
0.87 (0.40, 1.91)
rs6532796
0.99 (0.26, 3.73)
rs569557
1.14 (0.96, 1.36)
rs1149616
1.21 (0.32, 4.60)
rs498475
1.54 (0.30, 7.92)
rs7935829
2.55 (0.88, 7.35)
Multiple
GRS IVW
1.14 (0.97, 1.34)
GRS Egger
1.15 (0.94, 1.40)
GRS Weighted Median
1.14 (0.97, 1.34)
.1
1
10
Causal OR for CAD


IV Method
OR (95% CI)
Single
rs1260326
1.29 (0.52, 3.19)
rs569557
1.31 (1.07, 1.60)
rs10761784
1.45 (0.31, 6.72)
rs7935829
2.25 (0.74, 6.83)
rs6532796
2.63 (0.68, 10.15)
rs1149616
4.51 (0.89, 22.72)
Multiple
GRS IVW
1.37 (1.14, 1.65)
GRS Egger
1.27 (1.01, 1.59)
GRS Weighted Median
1.32 (1.09, 1.60)
.1
1
10
Causal OR for IS


IV Method
OR (95% CI)
Single
rs498475
0.07 (0.00, 4.28)
rs10761784
0.16 (0.00, 6.09)
rs1149616
0.30 (0.01, 10.39)
rs1260326
0.96 (0.14, 6.42)
rs6532796
1.10 (0.05, 23.87)
rs569557
1.31 (0.84, 2.07)
rs7935829
3.22 (0.24, 42.35)
Multiple
GRS IVW
1.22 (0.80, 1.85)
GRS Egger
1.48 (0.88, 2.49)
GRS Weighted Median
1.29 (0.84, 2.00)
.1
1
10
Causal OR for VTE

Causal effect estimates are shown as odds ratios (OR) amd 95% confidential intervals per every higher standard deviation change in FVII activity. Causal estimates based on single variant instrumental variables (IVs) are shown, as well as causal estimates based on the combination of these variants using inverse variance weighted (IVW) meta-analysis, Egger regression, and weighted median estimation.